ABSTRACT OF THE DISCLOSURE

A method and apparatus for automatically controlling the operation of a speech recognition system without requiring unusual or unnatural activity of the speaker by passively determining if received sound is speech of the user before activating the speech recognition system. A video camera and microphone are located in a hand-held device. The video camera records a video image of the speaker's face, i.e., of speech articulators of the user such as the lips and/or mouth. The recorded characteristics of the articulators are analyzed to identify the sound that the articulators would be expected to make, as in "lip reading". A microphone concurrently records the acoustic properties of received sound proximate the user. The recorded acoustic properties of the received sound are then compared to the characteristics of speech that would be expected to be generated by the recorded speech articulators to determine whether they match. If so, then the received sound is identified as having emanated from the user the speech recognition system is operated to perform speech recognition of the received sound.